AD-A079 273 ARMY PERSONNEL RESEARCH OFFICE WASHINGTON DC F/G 12/1
EVALUATION OF A MODIFIED OPTIMAL REGIONS ALOGORITHM FOR MANPOWE--ETC(U)
NOV 66 W D LARKIN

UNCLASSIFIED APRO-RM-66-9 NL

1 OF 1
AD
A079273

END
DATE
FILMED
2-80
DDC

LEVEL II

RECORD SET



Research Memorandum 66-9

# EVALUATION OF A MODIFIED OPTIMAL REGIONS ALGORITHM FOR MANPOWER SYSTEMS EXPERIMENTS



November 1966





APRO

U. S. ARMY PERSONNEL RESEARCH OFFICE

Army Project Number
2J024701A723

Computerized Manpower
Systems c-22

9

Research Memorandum 66-9

6

# EVALUATION OF A MODIFIED OPTIMAL REGIONS ALGORITHM FOR MANPOWER SYSTEMS EXPERIMENTS

10

By W. D. Larkin

Richard C. Sorenson, Task Leader

12 14

14 APRO-RM-66-9

Submitted by:
Cecil D. Johnson
Chief, Statistical Research
and Analysis Laboratory

Approved by:
J. E. Uhlaner
Director
USAPRO Laboratories

11

November 1966



040 650

slt

## FOREWORD

The Brogden-Weaver optimal regions algorithm was programmed in U. S. APRO for both a 12K IBM 1401 with tape transports and the much larger IBM 7094 computer system at the National Bureau of Standards. The 1401 version would achieve a near-optimal solution for a 3,000-by-75 matrix in a little more than 2 hours and, using more time, could solve a problem of twice that size--6,000 by 150. This program would read the operational data cards in the format prescribed by AR 611-259[1] and would adjust the assignee's predicted performance score to reflect his job preferences (if any). It would provide for the preclusion of assigning a man to a job for which he does not meet minimum prerequisites with respect to education, test scores, physical profile, security, etc., and would make assignments so as to maximize the average predicted performance of the individuals included in the assignment pool, given these restraints.

The Brogden-Weaver algorithm rapidly proceeds to a near-optimal feasible solution which from 95 to 99 percent of the assignees are optimally allocated and the remaining placed in jobs that are near optimal for these persons--jobs for which their performance scores are within a score interval of their highest adjusted score. However, this efficiency was accomplished using a range between two integers instead of a single integer to express the constraints (i.e., job quotas). Under this modification, the number that must be assigned to job X might be designated as any integer between 100 and 106 rather than as exactly 103. Ranges instead of single numbers were used only where actual management requirements were approximations in the first place, or where flexibility with respect to the number that would be acceptable was present.

While the Brogden-Weaver program was satisfactory for operational use, it was not satisfactory for use in the experimental simulation of manpower systems that are conducted in U. S. APRO. For these experiments, each replication must be constrained by exactly the same set of quota requirements and must have achieved an equivalent degree of optimality. The version of the optimal regions allocation method described and evaluated in the study reported here was attempting to obtain this required degree of exactness. As indicated in the Research Memorandum, the optimal regions method does not appear to be desirable for model sampling experiments. It should be noted that the Larkin modifications do not constitute as radical a departure from the classical Brogden-Dwyer approach as does the Brogden-Weaver algorithm.

Cecil D Johnson

CECIL D. JOHNSON
Chief
Statistical Research and
Analysis Laboratory



[1] Basic trainees available for advanced individual training, 23 September 1964.

# EVALUATION OF A MODIFIED OPTIMAL REGIONS ALGORITHM FOR MANPOWER SYSTEMS EXPERIMENTS

The basic data of the allocation problem are numbers, $a_{ij}$ which in practice represent aptitude scores of the ith man for the jth job. Alternatively, $a_{ij}$ could be an estimate of future achievement, a measure of efficiency or motivation, or any other index of the potential value of individual i to the organization if he were to perform job j. The solution of the allocation problem can be represented by a matrix $\{x_{ij}\}$ where $x_{ij} = 1$ if the ith man is assigned to the jth job, and $x_{ij} = 0$ otherwise. In its pure form, the problem is to allocate individuals to jobs such that the total utility $\sum_{i,j} a_{ij} x_{ij}$ of the assignments is the maximum that can accrue to the organization. This must be done under conditions that constrain the manpower supply and the job quotas. In the present investigation, these conditions were strict: the number of men available is exactly the number needed to fill the quotas, and the only acceptable solutions are those that assign every man to a job. Applications of the method are not expected to impose quite these conditions, e.g., the manpower pool may be larger or smaller than the number of available jobs and an incomplete allocation may be considered acceptable. Nonetheless, for research purposes, the imposition of strict conditions allows different allocation methods to be compared fairly.

## THE ALLOCATION SUM

In the present study, the measure of comparison was the allocation sum, a measure of the utility of a complete assignment, $\sum_{i,j} a_{ij} x_{ij}$. If we let $\underline{a_i} = \text{Min}\,\{a_{i,1}, a_{i,2}, \ldots , a_{i,M}\}$ and $\bar{a}_i = \text{Max}\,\{a_{i,1}, a_{i,2}, \ldots\ldots, a_{i,M}\}$, then the allocation sum is a number that lies between $\underline{A} = \sum_i \underline{a_i}$ and $\bar{A} = \sum_i \bar{a}_i$, which would be the allocation sums if each man were assigned to the job for which he has the least, or the greatest aptitude, respectively. The latter alternative is an upper bound on the performance of any allocation technique, but unless very special quotas are provided the assignment of men to the jobs for which they score highest is not possible. Thus, if the allocation sum for a possible assignment is denoted

by A, the difference $\bar{A}$ - A is an absolute measure of assignment adequacy. Similarly, if $A_1$ and $A_2$ are allocation sums for two assignment methods, the better method is the one that produces a larger allocation sum.

In addition to the quota restrictions, other contingencies may, in practice, affect an allocation solution. For example, a man who might otherwise be highly qualified for a job may be disqualified by a medical or security regulation, or by his personal aversion to the work. These extra conditions complicate, but do not change the essential character of the problem. Methods of dealing with such special cases are straightforward in principle, and later will become apparent.

## DEVELOPMENT OF THE OPTIMAL REGIONS METHOD

The optimal regions allocation method was proposed by Brogden (1946) and aspects of it were discussed by Dwyer (1954) and Brogden (1954). The method received its name from the formal similarity of the allocation problem to the geometric problem of partitioning an M-dimensional space into disjoint regions. For clarity, suppose that there are just two job assignment categories, $J_1$ and $J_2$ (i.e., M = 2, and that a sample of predicted performance scores is available, i.e., for each of N individuals a pair of numbers ($a_1$, $a_2$) indicates his expected performance for job assignments 1 and 2 respectively. Figure 1 shows a hypothetical scatterplot of these performance scores. Suppose that disjunctive criteria are desired for each assignment category, so that only those individuals having scores above a criterion are assigned to a job. These criteria are represented on the scatterplot as the vertical and horizontal lines $a_1 = c_1$ and $a_2 = c_2$. In effect, the adoption of these criteria is a rejection of all individuals in the rectangle between (0,0) and ($c_1$, $c_2$). These individuals are not eligible for either job. The assignments can be specified further by noting that the individuals in the shaded regions are eligible for only one of the two jobs. But the individuals with scores in the upper quadrant are in doubt. To separate these into assignment categories, the optimal regions method simply awards the assignments to the category with

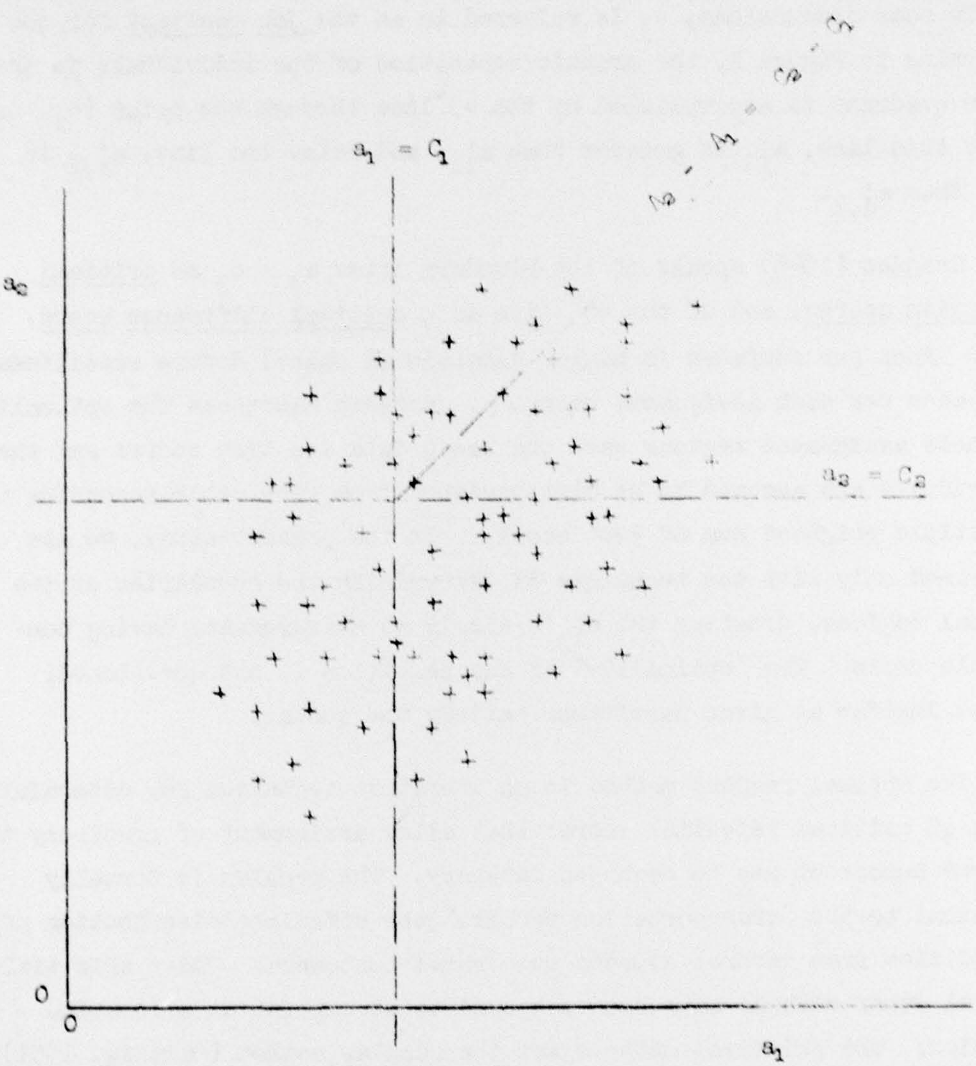


Figure 1. A two-dimensional score space divided into assignment regions by critical rejection scores.

the high[illegible] score relative to the criteria. The comparison of scores relative to criterion values is a fundamental part of the optimal regions method. Thus, each score, $a_{ij}$, is transformed to an adjusted score $a'_{ij} = a_{ij} - c_j$, where $c_j$ is the disjunctive criterion for job category j. In some discussions, $c_j$ is referred to as the job constant for job j. Returning to Figure 1, the graphic separation of the individuals in the upper quadrant is accomplished by the 45°line through the point $(c_1, c_2)$. Above this line, $a'_{i2}$ is greater than $a'_{i1}$, and below the line, $a'_{i,2}$ is less than $a'_{i,1}$.

Brogden (1946) speaks of the boundary lines $a_j = c_j$ as critical rejection scores, and of the 45° line as a critical difference score. These lines (or surfaces in higher dimensional space) define rectilinear subspaces for each assignment category. Brogden discusses the optimality of these assignment regions when the basic data are test scores and the individuals are assumed to be distinguished from each other according to a multiple weighted sum of test scores. In the present study, we are concerned only with the technique of determining the boundaries of the optimal regions, treating the $a_{ij}$'s simply as measurements having comparable units. The "optimality" of any partition is not questioned, except insofar as given partitions satisfy the quotas.

The optimal regions method is an iterative technique for determining a set of critical rejection scores that allow assignment of precisely the desired number of men to each job category. The problem is formally identical to the "transportation problem" the efficient distribution of commodities from several sources to several customers. Under this title, several other methods have been proposed to accomplish an allocation solution. The principal methods are the simplex method (Dantzig, 1951), the Hungarian method (Kuhn, 1955), and the method of Ford and Fulkerson (1956). Each of these is also iterative, although the steps are somewhat different. The otpimal regions method appears to have a potential advantage over other methods when the matrix of scores is very large, inasmuch as it is not necessary to deal with the entire score matrix at once. In particular, if the allocation can be accomplished row by row, the size of the

problem will not be seriously limited by the storage capacity of a computer. The principal objective of the present study was to devise a row by row processing technique that furnishes a set of optimal regions.

The problem to be solved can again be illustrated graphically. Returning to Figure 1, a different choice of critical rejection scores could be made which would effect two changes in the allocation pattern: 1) the new rejection scores would raise or lower the total number of assignments, and 2) the relative locations of the new rejection scores could induce a change in the proportion of men assigned to each category. The purpose of the iterative allocation procedure is to construct a sequence of sets of critical rejection scores that converges to a set that allows complete allocation, with the desired number of men in each category. The technique proposed by Brogden (1946) begins with the marginal frequency distributions of $a_{ij}$ scores for each job. The initial rejection scores are estimated from these distributions by "accepting for consideration" the $Q_j$ men who score highest in each distribution, where $Q_j$ is the pre-set quota. In other words, each initial rejection score is determined so that the total number of men scoring above it in the jth marginal distribution is (approximately) $Q_j$. Next, wherever an individual's score exceeds two or more rejection scores, the adjusted score $a_{ij} - c_j$ is computed and the individual is tentatively assigned to the category for which he has the higher score. The number of individuals eligible for more than one job depends on the correlation between the scores; however, even with uncorrelated scores, a portion of the initially "accepted" men will score higher than two or more rejection criteria. In general, therefore, fewer than $Q_j$ assignments will actually be made to job category j, and the initial estimate of the critical rejection score will result in only partial satisfaction of the quota. Thus, after the first estimates, each job category remains in deficit and new estimates are required for which these deficits serve as quotas. The second, and each succeeding iteration, is performed exactly as the first, except that the inspection for multiple assignments must cover all previous iterations. In principle, the process continues until all men are assigned and none are assigned to more than one job.

DEALING WITH TIED SCORES

In practice, the mechanics of the process are not difficult to realize in a computer program, but serious difficulties are encountered in bringing the procedure to a satisfactory solution. The principal source of trouble is the lack of an efficient method to deal with adjusted scores that are tied. This problem arises whenever the scores lack precision. Aptitude test scores used in the present investigation were simulated two-digit numbers between 01 and 99, normally distributed about a mean of 50. (Occasional outliers, greater than 99, were included.) The use of simulated scores allowed complete control of the statistical properties of the score matrix, and at the same time provided a reasonable representation of Army standard scores on the Army Classification Battery. In several of the matrices used, the correlation coefficients between pairs of tests were set at 0.7. Given these rather highly correlated two-digit numbers, it was not unusual to encounter tied scores among 15% or 20% of the men. In an initial version of the program, these tied scores were treated in sequence as they arose with untied scores. On several 300 x 7 score matrices, the program was unable to assign more than about 85% of the men. This difficulty led to experiments with corrective procedures that improved upon, but did not complete the optimal regions procedure. Some of these corrections became fully as detailed and time-consuming as the basic method, but none were satisfactory. Some of them will be described briefly before the modified optimal regions method is discussed.

The optimal regions method does not deal explicitly with tied scores. Nevertheless, demonstrations of the method often make use of two notations for assignments: one to indicate that an individual is assigned unambiguously to a job, and another to indicate that the assignment is tied with one or more others. (See, for example, the notation used by Dwyer, 1954). If the assignment problem is relatively small, it soon becomes apparent that, by breaking the ambiguities in the proper way, the quotas can be filled. The difficulty with a large assignment problem is that this tie-breaking procedure must be translated into an exact method. The sequential treatment of ties implied by the basic optimal regions algorithm is not exact, because the ties are broken one-at-a-time, and the total pattern

of ties does not become apparent until the last man has been assigned. To make the procedure exact, the tied scores could be stored in a submatrix that could be processed after all untied assignments were complete. One possibility, which was not tried in the study, is to treat the submatrix of ties with a Hungarian or Simplex procedure. In practice, this might prove to be an efficient procedure.

As a substitute for a completely separate treatment of ties, two alternatives were investigated:

a) Ordering the score matrix so that men were allocated in order of decreasing variability in their scores.

b) Allocating men with tied scores only after all other men have been processed.

The first technique did not noticeably improve the method: about 6% of the men could not be assigned. When combined with the first, the second technique further reduced the pool of unassigned men, but still failed to reach a solution. Both techniques were designed to increase the amount of quota information available at the time the ties were broken.

The basic difficulty in the optimal regions paradigm is that a universe of continuous scores in which no ties are possible is assumed, whereas in practice the scores are always discrete. To avoid ties, it is enough to be able to construct a complete ordering of the scores in the score matrix. This, however, is not usually possible unless the scores are known with great precision. If the aptitude area scores were four-digit numbers instead of two, few, if any, ties woud be expected in a pool of 300 men, but even four digit scores might be too imprecise if there were 1000 men to allocate. The effect of the tied (and nearly tied) scores is that, as the method approaches a solution, the adjustments it can make to shift men from one assignment to another are overly coarse. The minimum adjustment is the addition of an integer to one of the rejection criteria. If two or more individuals are tied in one assignment category, this minimal adjustment can assign all of them to the category, or none of them, but not some without the others.

Several attempts were made to bypass the problem of tied scores with special adjustments of the computed rejection scores. One of these was put into effect after the basic iterations began to produce rejection scores in cycles. As soon as the cyclic condition was discovered, the score information was slightly degraded by treating scores one unit removed from the rejection scores as equivalent to the rejection scores. The effect of this loss of resolution is that the pool of tied scores is enlarged, and presumably the quotas are therefore more easily met. This attempt to lessen the problem of ties could not be completely evaluated, because the speed of convergence to a possible solution was excessively slow.

No method for deciding ties has been satisfactory, because the row by row treatment of the score matrix does not permit accumulation of information sufficient to guarantee correct tie decisions, even when a set of correct decisions exists. The method used by Dywer succeeded because the entire matrix pattern of tied scores was treated as an essentially separate problem. But the use of an ancillary algorithm to deal with ties is not the only alternative: the optimal regions method can be modified to allocate in sections, as described below.

## MODIFIED OPTIMAL REGIONS METHOD

The central idea of the modified optimal regions technique is to accept a series of partial solutions rather than demand a single complete solution. With minor changes, each partial solution is obtained just as in Brogden's (1946) original description. When no improvement can be made in a partial solution, the assignments completed are fixed, and the procedure is re-started with the remaining pool of unassigned men. This series of partial assignments does not preclude the possibility that a small group of individuals with tied scores will remain incapable of assignment, but such a group is likely to be so small relative to the total personnel pool that, however they are assigned, the allocation sum will not be greatly affected.

The principal steps are as follows:

1. Read the score matrix, and establish a frequency table of scores for each job. Also, determine $\tilde{A}$, the maximal allocation sum.

2. Determine initial rejection scores by counting down the frequency table to the $Q_j$ th man, and pick the lowest rejection score that would not result in over-filling the quota if all men with higher scores were assigned. Clear frequency table.

3. Read score matrix again, and

   (a) Subtract critical rejection scores from each raw score.

   (b) Enter all adjusted scores in the score frequency tables.

   (c) Assign each man to the job for which his adjusted score is highest. If two or more adjusted scores are tied, assign the man to the job for which his raw score is highest. If the raw scores are also tied, make the assignment to the first even-numbered job if the man number is even, and to the first odd-numbered job if the man number is odd. If all adjusted scores are negative, no assignment can be made.

4. Determine which job has the greatest percentage deficiency.

   (a) Lower the rejection score for this job, adding as many men as possible without overfilling the quota. If the rejection score cannot be made lower, go to the job with the next highest percentage deficiency.

   (b) If a rejection score was lowered, read score matrix again and make new assignments.

   (c) If quotas are all satisfied, go to step 5.

   (d) If quotas are not satisfied but no rejection score could be lowered without over-filling a quota, go to step 6.

5. Punch out assignments, allocation sum, total number of iterations, and stop.

6. Punch out assignments, allocation sum, number of iterations, and the matrix of unassigned men. Go to step 1.

This procedure is initially slower than the method proposed by Brogden, because, on each iteration, only one rejection score is lowered. This precaution is taken to prevent overfilling of job quotas--a possibility when all rejection scores are lowered at once. The slowness of the convergence is partly compensated by the finer adjustments possible when the program approaches a solution.

Three typical results, using the modified optimal regions program, are given as follows:

| | A | $\bar{A}$ | No. of Iterations |
|---|---|---|---|
| Problem I (300 men) | 19579 | 19811 | 133 |
| Problem II (100 men) | 6589 | 6624 | 89 |
| Problem III (160 men) | | | |
| Optimal Region | 9084 | 9113 | 78 |
| Ford-Fulkerson Program | 9106 | --- | --- |

Results indicate that the modification of the optimal regions method apparently does not seriously reduce its ability to produce a good allocation. In Problem III, the allocation sum compares favorably with the one produced by the Ford-Fulkerson program. The results also suggest that it may not be necessary to deal with large personnel pools in order to achieve a satisfactory assignment: optimal assignments of sub-populations may accomplish nearly the same allocation sum as assignment of entire population.

## REFERENCES


Brogden, H. E. An approach to the problem of differential prediction. Psychometrika, 11, 1946, 139-154.

Brogden, H. E. A simple proof of a personnel classification theorem. Psychometrika, 19, 1954, 205-208.

Dwyer, P. S. Solution of the personnel classification problem with the method of optimal regions. Psychometrika, 19, 1954, 11-25.

Dantzig, G. B. Application of the simplex method to a transportation problem in "Activity analysis of production and allocation." Cowles Commission Monograph, 13, p. 359, J. Wiley and Sons, New York, 1951.

Kuhn, H. W. The Hungarian method for the assignment problem., Naval Research Logistics Quarterly, 2, 1955, p. 83.

Ford, L. R. and Fulkerson, D. R. Solving the transportation problem. Management Science, 3, 1956, p. 24.

Ford, L. R. and Fulkerson, D. R. A primal-dual algorithm for the capacitated Hitchcook problem. RAND Research Memorandum, p. 827, Santa Monica, 1956.